butyl (S) chloro-5-hydroxy-3-oxohexanoate (*e.g.*, U.S. Patent No. 6,645,746 and WO 01/40450), reduction of pyrrolotriazine-based compounds (*e.g.*, U.S. application No. 2006/0286646); reduction of substituted acetophenones (*e.g.*, U.S. Patent No. 6,800,477); and reduction of ketothiolanes (WO 2005/054491). In another approach, as demonstrated herein, the ketoreduction can be carried out in the presence of an alcohol, such as isopropanol, to provide a substrate for the reverse reaction (alcohol dehydrogenation). In this manner, the NADH/NADPH consumed in the ketoreduction reaction is regenerated by the reverse, oxidative reaction.

[0006] U.S. Patent No. 7,083,973 discloses a stereoselective process for the preparation of (*3S,2R*)-1-halo-2-hydroxy-3-(protected)amino-4-substituted butanes by the reduction of the corresponding keto group containing compounds using certain species of *Rhodococcus* and *Brevibacterium*. The '973 patent discloses that only selected species of *Rhodococcus* and *Brevibacterium* catalyze the reduction to form the desired (*3S,2R*)-1-halo-2-hydroxy-3-(protected)amino-4-substituted butanes in high quantitative and enantiomeric yield. The '973 patent discloses that 10 mL of cell extract from 150 g of *Rhodococcus erythropolis* ATCC 4277 cells loaded with 10 mg of (1*S*)-[N-(1-benzyl-2-oxo-3-chloro)propyl]carbamic acid t-butyl ester substrate, glucose dehydrogenase (35 units), 0.7 mM NAD^+ and 200 mg of glucose (reaction carried out at pH 6.0, 150 RPM agitation and 30 °C) results in (1S,2R)-[N-(1-benzyl-2-hydroxy-3-chloro)propyl]carbamic acid t-butyl ester product in 95% yield with >98% diastereomeric purity.

[0007] Accordingly, isolated KRED polypeptides capable of stereoselective conversion of α-halo-ketones to chiral halo-alcohols in high-yield and high diastereomeric purity would be desirable. Also, improved processes for using KRED polypeptides to carry out large-scale preparation of chiral halo-alcohols would be desirable.

4. SUMMARY

[0008] The present disclosure provides ketoreductase polypeptides capable of stereoselectively converting an α-halo-ketone to a chiral halo-alcohol, and methods for using these polypeptides in synthetic processes for making chemical compounds such as intermediates in the production of active pharmaceutical ingredients, such as the antiretroviral drug, atazanavir.

[0008a] In one aspect, there is provided an engineered ketoreductase polypeptide capable of converting (S)-tert-butyl 4-chloro-3-oxo-1-phenylbutan-2-ylcarbamate to tert-butyl (2S,3R)-4-chloro-3-hydroxy-1-phenylbutan-2-ylcarbamate with a relative activity at least about 1.2-fold that of SEQ ID NO:2, wherein the polypeptide comprises an amino acid sequence at least about 70% identical to SEQ ID NO:2 and includes one or more of the following features:

residue corresponding to amino acid 2 of SEQ ID NO:2 is an aliphatic or nonpolar amino acid selected from alanine, leucine, valine, isoleucine, glycine and methionine;

residue corresponding to amino acid 28 of SEQ ID NO:2 is an aliphatic or nonpolar amino acid selected from alanine, leucine, valine, isoleucine, glycine and methionine;
residue corresponding to amino acid 34 of SEQ ID NO:2 is a polar amino acid selected from asparagine, glutamine, serine, and threonine;
residue corresponding to amino acid 47 of SEQ ID NO:2 is an aliphatic or nonpolar amino acid selected from alanine, leucine, valine, isoleucine, glycine and methionine;
residue corresponding to amino acid 50 of SEQ ID NO:2 is a basic amino acid selected from lysine and arginine;
residue corresponding to amino acid 81 of SEQ ID NO:2 is a polar amino acid selected from among asparagine, glutamine, serine, and threonine;
residue corresponding to amino acid 90 of SEQ ID NO:2 is an aliphatic or nonpolar amino acid selected from alanine, leucine, valine, isoleucine, glycine and methionine;
residue corresponding to amino acid 91 of SEQ ID NO:2 is an aliphatic or nonpolar amino acid selected from alanine, leucine, valine, isoleucine, glycine and methionine, an aromatic amino acid selected from tyrosine, tryptophan, and phenylalanine, or a basic amino acid selected from lysine and arginine;
residue corresponding to amino acid 94 of SEQ ID NO:2 is the basic amino acid, arginine;
residue corresponding to amino acid 112 of SEQ ID NO:2 is an aromatic amino acid selected from tyrosine, tryptophan, and phenylalanine;
residue corresponding to amino acid 117 of SEQ ID NO:2 is an acidic amino acid selected from aspartic acid and glutamic acid;
residue corresponding to amino acid 143 of SEQ ID NO:2 is a basic amino acid selected from lysine and arginine;
residue corresponding to amino acid 144 of SEQ ID NO:2 is a polar amino acid selected from asparagine, glutamine, serine, and threonine;
residue corresponding to amino acid 145 of SEQ ID NO:2 is a nonpolar amino acid selected from alanine, leucine, valine, isoleucine, and methionine or an aliphatic amino acid selected from alanine, leucine, valine, isoleucine;
residue corresponding to amino acid 148 of SEQ ID NO:2 is a constrained amino acid selected from proline and histidine;
residue corresponding to amino acid 150 of SEQ ID NO:2 is a nonpolar or aliphatic amino acid selected from leucine, valine, isoleucine, glycine and methionine;
residue corresponding to amino acid 152 of SEQ ID NO:2 is a nonpolar or aliphatic amino acid selected from alanine, leucine, valine, isoleucine, glycine and methionine;
residue corresponding to amino acid 153 of SEQ ID NO:2 is a nonpolar or aliphatic amino acid selected from alanine, leucine, valine, isoleucine, glycine and methionine, or a constrained amino acid selected from histidine and proline;

residue corresponding to amino acid 158 of SEQ ID NO:2 is a polar amino acid selected from asparagine, glutamine, and serine;

residue corresponding to amino acid 190 of SEQ ID NO:2 is a nonpolar or an aliphatic amino acid selected from alanine, valine, leucine, isoleucine, glycine, and methionine;

residue corresponding to amino acid 198 of SEQ ID NO:2 is a polar amino acid selected from asparagine, glutamine, and threonine;

residue corresponding to amino acid 199 of SEQ ID NO:2 is an aliphatic or nonpolar amino acid selected from alanine, leucine, valine, glycine, and methionine, or a polar amino acid selected from asparagine, glutamine, serine, and threonine;

residue corresponding to amino acid 200 of SEQ ID NO:2 is a nonpolar amino acid selected from alanine, leucine, valine, isoleucine, and glycine;

residue corresponding to amino acid 217 of SEQ ID NO:2 is a polar amino acid selected from asparagine, glutamine, serine and threonine;

residue corresponding to amino acid 225 of SEQ ID NO:2 is a nonpolar amino acid selected from valine, leucine, glycine, and methionine;

residue corresponding to amino acid 231 of SEQ ID NO:2 is an aromatic amino acid selected from tyrosine, tryptophan, and phenylalanine;

residue corresponding to amino acid 232 of SEQ ID NO:2 is a nonpolar amino acid selected from leucine, isoleucine, valine, glycine, and methionine;

residue corresponding to amino acid 233 of SEQ ID NO:2 is a polar amino acid selected from asparagine, glutamine, serine, and threonine;

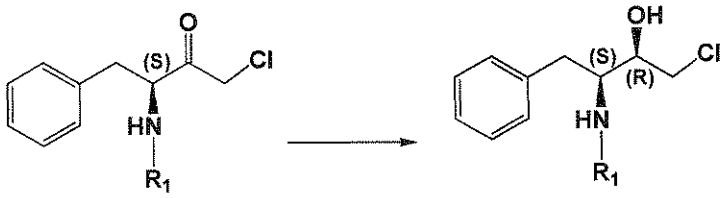
residue corresponding to amino acid 244 of SEQ ID NO:2 is a nonpolar amino acid selected from alanine, leucine, isoleucine, valine, glycine, and methionine;

residue corresponding to amino acid 260 of SEQ ID NO:2 is an aromatic amino acid selected from tyrosine and tryptophan; and

residue corresponding to amino acid 261 of SEQ ID NO:2 is a polar amino acid selected from asparagine, glutamine, and threonine.

[0009] In certain embodiments, the disclosure provides ketoreductase polypeptides capable of converting N-protected (S)-3-amino-1-chloro-4-phenylbutan-2-one compounds of Formula (I) (in which R^1 is a protecting group) to the corresponding stereoisomeric alcohol N-protected (2R,3S)-3-amino-1-chloro-4-phenylbutan-2-ol product of Formula (II), as depicted in Scheme 1 below:

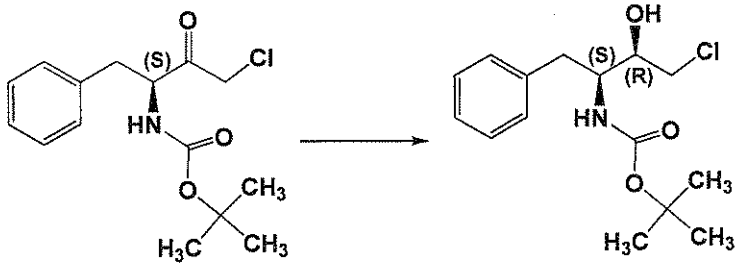
Scheme 1



Formula (I) Formula (II)

[0010] In particular embodiments, the present disclosure provides ketoreductase polypeptides capable of converting chloroketone compound (**1**) ((S)-tert-butyl 4-chloro-3-oxo-1-phenylbutan-2-ylcarbamate) to the corresponding alcohol, compound (**2**) (tert-butyl (2S,3R)-4-chloro-3-hydroxy-1-phenylbutan-2-ylcarbamate), as depicted in Scheme 2 below:

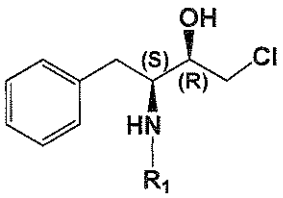
Scheme 2



(1) (2)

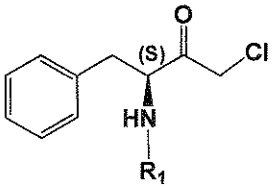
[0011] In certain embodiments, the present disclosure provides ketoreductase polypeptides capable of converting a reaction mixture comprising an initial concentration of at least 10 g/L compound (**1**) to compound (**2**) with a conversion rate of at least 70% in 24 hours. In certain embodiments, the concentration of polypeptide capable of carrying out this conversion is 5 g/L, 2 g/L, 1 g/L, or less. In certain embodiments, the polypeptides are capable of a conversion rate of at least 80%, 85%, 90%, 95%, 98%, 99%, or more in 24 hours, or even less time. In certain embodiments, the polypeptides are capable of converting compound (**1**) to compound (**2**) in a diastereomeric excess of greater than about 95%, greater than about 97%, or greater than about 99%. In certain embodiments, the polypeptides are capable of the above conversion rates with a reaction mixture comprising an initial concentration of compound (**1**) of at least 20 g/L, 40 g/L, 60 g/L, 80 g/L, 100 g/L, 150 g/L, 200 g/L, or even more.

[0011a] In one aspect, there is provided a process for preparing a compound of Formula (II) in greater than about 95% diastereomeric excess, said process comprising:



Formula (II)

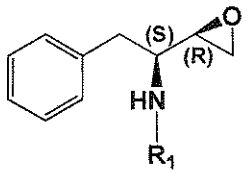
contacting a compound of Formula (I)



Formula (I)

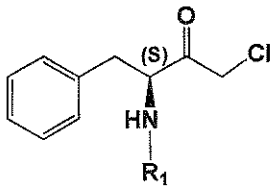
with an engineered ketoreductase polypeptide as defined above under suitable reaction conditions.

[0011b] In another aspect, there is provided a process for preparing a compound of Formula (III) in greater than about 95% diastereomeric excess, said process comprising:



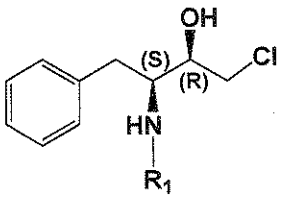
Formula (III)

(a) contacting a compound of Formula (I) with a polypeptide as defined above under suitable reaction conditions,



Formula (I)

thereby forming a reaction mixture comprising a compound of Formula (II)



Formula (II) ;

(b) extracting the reaction mixture with an organic solvent; and

(c) contacting the organic solvent extract with a base.

[0012] In certain embodiments, the present disclosure provides a method for converting a compound of Formula (**I**) to a compound of Formula (**III**) (see Scheme 3), wherein R^1 is as described above, comprising converting the compound of Formula (**I**) to a compound of Formula (**II**) using a ketoreductase of the present disclosure and then contacting the compound of Formula (**II**) with base to provide the compound of Formula (**III**).

Scheme 3

disclosed herein under reaction conditions suitable for reducing or converting the substrate to the product.

[0017] In some embodiments of the above methods, the substrate is reduced to the product in greater than about 95%, greater than about 97%, or greater than about 99% diastereomeric excess, wherein the ketoreductase polypeptide comprises a sequence that corresponds to SEQ ID NO: 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56 and 58.

[0018] In certain embodiments of the above methods, at least about 95% of the substrate is converted to the product in less than about 24 hours when carried out with greater than about 100 g/L of substrate and less than about 5 g/L of the polypeptide. In certain embodiments, the polypeptide capable of carrying out the method comprises an amino acid sequence corresponding to SEQ ID NO: 6, 50, 52, and 56. In some embodiments of the above methods, at least about 95% of the substrate is converted to the product in less than about 30 hours when carried out with greater than about 150 g/L of substrate and less than about 1 g/L of the polypeptide, wherein the polypeptide comprises an amino acid sequence corresponding to SEQ ID NO: 6, 50, 52, and 56.

[0019] In one ~~aspect~~embodiment, the ketoreductase polypeptides described herein have amino acid sequences with one or more amino acid differences as compared to a wild-type ketoreductase or as compared to an engineered ketoreductase. The one or more amino acid differences result in at least one improved property of the enzyme for a defined substrate. Generally, the ketoreductase polypeptides described herein are engineered ketoreductase polypeptides having one or more improved properties as compared to the naturally-occurring wild-type ketoreductase enzymes obtained from *Novosphingobium aromaticivorans* ("*N. aromaticivorans*"; SEQ ID NO:2). Improvements in an enzyme property of the engineered ketoreductase polypeptides include increases in enzyme activity, stereoselectivity, sterospecificity, thermostability, solvent stability, tolerance to increased levels of substrate, and tolerance to increased levels product.

[0020] In some embodiments, the ketoreductase polypeptides of the ~~invention~~disclosure are improved as compared to SEQ ID NO:2 with respect to their rate of enzymatic activity, i.e., the conversion rate for reducing N-protected (S)-3-amino-1-chloro-4-phenylbutan-2-one ("the substrate") (e.g. compound (**1**) where the protecting group is a BOC moiety) to N-protected (2R,3S)-3-amino-1-chloro-4-phenylbutan-2-ol ("the product") (*e.g.,* compound (**2**)). In some embodiments, the engineered ketoreductase polypeptides are capable of converting the substrate to the product at a conversion rate that is at least at least 1.1-times, 1.2-times, 1.3-times, 1.5-times, 2-times, 3-times, or more than 3-times the rate exhibited by the enzyme of SEQ ID NO: 2 under comparable assay conditions.

[0021] In some embodiments, such ketoreductase polypeptides are also capable of converting the N-protected (S)-3-amino-1-chloro-4-phenylbutan-2-one ("the substrate") (e.g., compound (**1**) where the protecting group is a BOC moiety) to N-protected (2R,3S)-3-amino-1-chloro-4-phenylbutan-2-ol

that have this capability include, but are not limited to, polypeptides which comprise amino acid sequences corresponding to SEQ ID NO: 6, 50, 52, and 56.

[0026] In some embodiments, the ketoreductase polypeptide is highly stereoselective, wherein the polypeptide can reduce the substrate to the product in greater than about 99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8% or 99.9% diastereomeric excess. Exemplary ketoreductase polypeptides with high stereoselectivity include, but are not limited to, the polypeptides comprising the amino acid sequences corresponding to SEQ ID NO: 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56 and 58.

[0026a] In another aspect, there is provided an engineered ketoreductase polypeptide capable of oxidizing isopropanol (IPA) to acetone with an activity at least 2-fold greater than the reference polypeptide of SEQ ID NO: 2, wherein the polypeptide comprises an amino acid sequence having at least 70% identity to a sequence selected from SEQ ID NO: 6, 56, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, and 80.

[0027] In some embodiments, the ketoreductase polypeptide has the improved property of increased activity with a secondary alcohol for cofactor regeneration. In some embodiments, the ketoreductase polypeptide oxidizes isopropanol (IPA) to acetone with an activity at least 2-fold, 2.5-fold, 5-fold, 10-fold, 15-fold, or even greater, relative to the reference polypeptide of SEQ ID NO: 2. Exemplary ketoreductase polypeptides exhibiting the improved property of increased activity with IPA include, but are not limited to, the polypeptides comprising the amino acid sequence corresponding to SEQ ID NO: 6, 56, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, and 80.

[0028] In certain embodiments, a ketoreductase polypeptide of the present disclosure is at least about 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 70%, 71%, 72%, 73%,74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to the reference sequence of SEQ ID NO:2, and has, at the position corresponding to the indicated position of SEQ ID NO:2, at least one of the following amino acid differences: the amino acid at position **2** is an aliphatic or nonpolar amino acid selected from among alanine, leucine, valine, isoleucine, glycine and methionine; the amino acid at position **28** is an aliphatic or nonpolar amino acid selected from among alanine, leucine, valine, isoleucine, glycine and methionine; the amino acid at position **34** is a polar amino acid selected from among asparagine, glutamine, serine, and threonine; the amino acid at position **47** is an aliphatic or nonpolar amino acid selected from among alanine, leucine, valine, isoleucine, glycine and methionine; the amino acid at position **50** is a basic amino acid selected from lysine and arginine; the amino acid at position **81** is a polar amino acid selected from among asparagine, glutamine, serine, and threonine; the amino acid at position **90** is an aliphatic or nonpolar amino acid selected from among alanine, leucine, valine, isoleucine, glycine and methionine; the amino acid at position **91** is either an aliphatic or nonpolar amino acid selected from among alanine, leucine, valine, isoleucine, glycine and methionine, or the amino acid at position **91** is an aromatic amino acid selected from among tyrosine, tryptophan, and phenylalanine, or the amino acid at position **91** basic amino acid selected from among lysine and arginine; the amino acid at position **94** is the basic amino acid, arginine; the amino acid at position **112** is an aromatic amino acid selected from among tyrosine, tryptophan, and phenylalanine; the amino acid at position **117** is an acidic

the proline residue at position **231** is replaced with phenylalanine; the alanine residue at position **232** is replaced with valine; the glutamic acid residue at position **233** is replaced with glutamine; the aspartic acid residue at position **244** is replaced with glycine; the phenylalanine residue at position **260** is replaced with tyrosine; and the serine residue at position **261** is replaced with asparagine.

[0031] In some embodiments, the ketoreductase polypeptides can have, in addition to the above, one or more modifications (*i.e.*, residue differences) as compared to the reference amino acid sequence or as compared to any of SEQ ID NO:4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, or 80. These modifications can be amino acid insertions, deletions, substitutions, or any combination of such changes. In some embodiments, the amino acid sequence differences can comprise non-conservative, conservative, as well as a combination of non-conservative and conservative amino acid substitutions. In some embodiments, these ketoreductase polypeptides can have optionally from about 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-11, 1-12, 1-14, 1-15, 1-16, 1-18, 1-20, 1-22, 1-24, 1-25, 1-30, 1-35 or about 1-40 mutations at other amino acid residues. In some embodiments, the number of modifications can be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15, 16, 18, 20, 22, 24, 26, 30, 35 or about 40 other amino acid residues.

[0032] In some embodiments, an improved ketoreductase comprises an amino acid sequence that is at least about 70%, 71%, 72%, 73%,74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to the amino acid sequence corresponding to SEQ ID NO: 2, wherein the improved ketoreductase polypeptide amino acid sequence includes any one set of the specified amino acid substitution combinations presented in Tables 2 or 3. In some embodiments, these ketoreductase polypeptides can have mutations at other amino acid residues.

[0032a] In one aspect, there is provided a composition comprising the polypeptide as defined above.

[0033] In ~~another aspect~~one embodiment, the present disclosure provides polynucleotides encoding the ketoreductase polypeptides described herein, and polynucleotides that hybridize to such polynucleotides under highly stringent conditions. In one aspect, there is provided a polynucleotide encoding a polypeptide as defined above. The polynucleotide can include promoters and other regulatory elements useful for expression of the encoded engineered ketoreductase, and can utilize codons optimized for specific desired expression systems.

[0034] In some embodiments, the disclosure provides polynucleotides encoding ketoreductase polypeptides having at least the following amino acid sequence as compared to the amino acid sequence of SEQ ID NO:2, and further comprising at least one acid substitution selected from the group consisting of: the proline residue at position 2 is replaced with leucine; the valine residue at position 28 is replaced with alanine; the alanine residue at position 34 is replaced with serine; the alanine residue at position 47 is replaced with valine; the glutamic acid residue at position 50 is replaced with lysine; the aspartic acid residue at position 81 is replaced with asparagine; the serine

residue at position 90 is replaced with valine; the isoleucine residue at position 91 is replaced with an amino acid selected from among leucine, tryptophan, arginine, and lysine; the lysine residue at position 94 is replaced with arginine; the aspartic acid residue at position 112 is replaced with tyrosine; the glycine residue at position 117 is replaced with aspartic acid; the serine residue at position 143 is replaced with arginine; the valine residue at position 144 is replaced with an amino acid selected from among cysteine and threonine; the glycine residue at position 145 is replaced with an amino acid selected from among alanine and valine; the arginine residue at position 148 is replaced with histidine; the alanine residue at position 150 is replaced with an amino acid selected from among glycine, isoleucine, serine, and tryptophan; the phenylalanine residue at position 152 is replaced with leucine; the asparagine residue at position 153 is replaced with an amino acid selected from among glycine, valine, and histidine; the threonine residue at position 158 is replaced with serine; the glycine residue at position 190 is replaced with an amino acid selected from among alanine, proline, glutamine, and valine; ; the serine residue at position 198 is replaced with asparagine; the isoleucine residue at position 199 is replaced with an amino acid selected from among glycine, methionine, leucine, and asparagine; the methionine residue at position 200 is replaced with isoleucine; the valine residue at position 204 is replaced with phenylalanine; the alanine residue at position 217 is replaced with threonine; the isoleucine residue at position 225 is replaced with valine; the proline residue at position 231 is replaced with phenylalanine; the alanine residue at position 232 is replaced with valine; the glutamic acid residue at position 233 is replaced with glutamine; the aspartic acid residue at position 244 is replaced with glycine; the phenylalanine residue at position 260 is replaced with tyrosine; and the serine residue at position 261 is replaced with asparagine. Exemplary polynucleotides include, but are not limited to, a polynucleotide sequence of SEQ ID NO: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, and 79.

[0034a] In one aspect, there is provided an expression vector comprising the polynucleotide as defined above operably linked to control sequences capable of directing expression of the encoded polypeptide in a host cell.

[0035] In some embodiments, the present disclosure provides host cells comprising the polynucleotides encoding the ketoreductase polypeptides as described herein and/or expression vectors comprising these polynucleotides.

[0035a] In one aspect, there is provided a host cell comprising the expression vector as defined above.

[0035b] The host cells may be *N. aromaticivorans* or they may be a different organism, and as *E. coli*. The host cells can be used for the expression and isolation of the engineered ketoreductase enzymes described herein, or, alternatively, they can be used directly for the conversion of the substrate to the stereoisomeric product. Accordingly, in some embodiments, the engineered ketoreductase polypeptides disclosed herein can be prepared by standard methods comprising culturing a host cell containing an expression vector comprising a polynucleotide encoding the polypeptide and isolating the polypeptide from the host cell.

[0035c] In another aspect, there is provided a method for preparing an engineered ketoreductase polypeptide as defined above comprising culturing a host cell as defined above and isolating the polypeptide from the host cell.

[0036] Whether carrying out the method with whole cells, cell extracts or purified ketoreductase enzymes, a single ketoreductase enzyme may be used or, alternatively, mixtures of two or more ketoreductase enzymes may be used.